2641

DOCKET NO. BABU 1-10-42

PATENT

#3
660803
S.I

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Shivnath Babu, et al.

Serial No.: 10/033,199

Filed: December 28, 2001

For: SYSTEM AND METHOD FOR COMPRESSING A DATA TABLE USING MODELS

Group: 2641

Examiner: N/A

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Alexandria, VA 22313, on
May 20, 2003 (Date)
Stephanie Pitt
(Printed or typed name of person signing the certificate)
Stephanie Pitt
(Signature of the person signing the certificate)

Sir:

LETTER TO OFFICIAL DRAFTSMAN

Transmitted herewith are eight sheets of formal drawings to be substituted for the informal drawings initially filed in the above-identified application for patent.

Respectfully submitted,

HITT GAINES & BOISBRUN, P.C.

David H. Hitt
Registration No. 33,182

Date: MAY 19, 2003

P.O. Box 832570
Richardson, Texas 75083
(972) 480-8800

FIG. 1

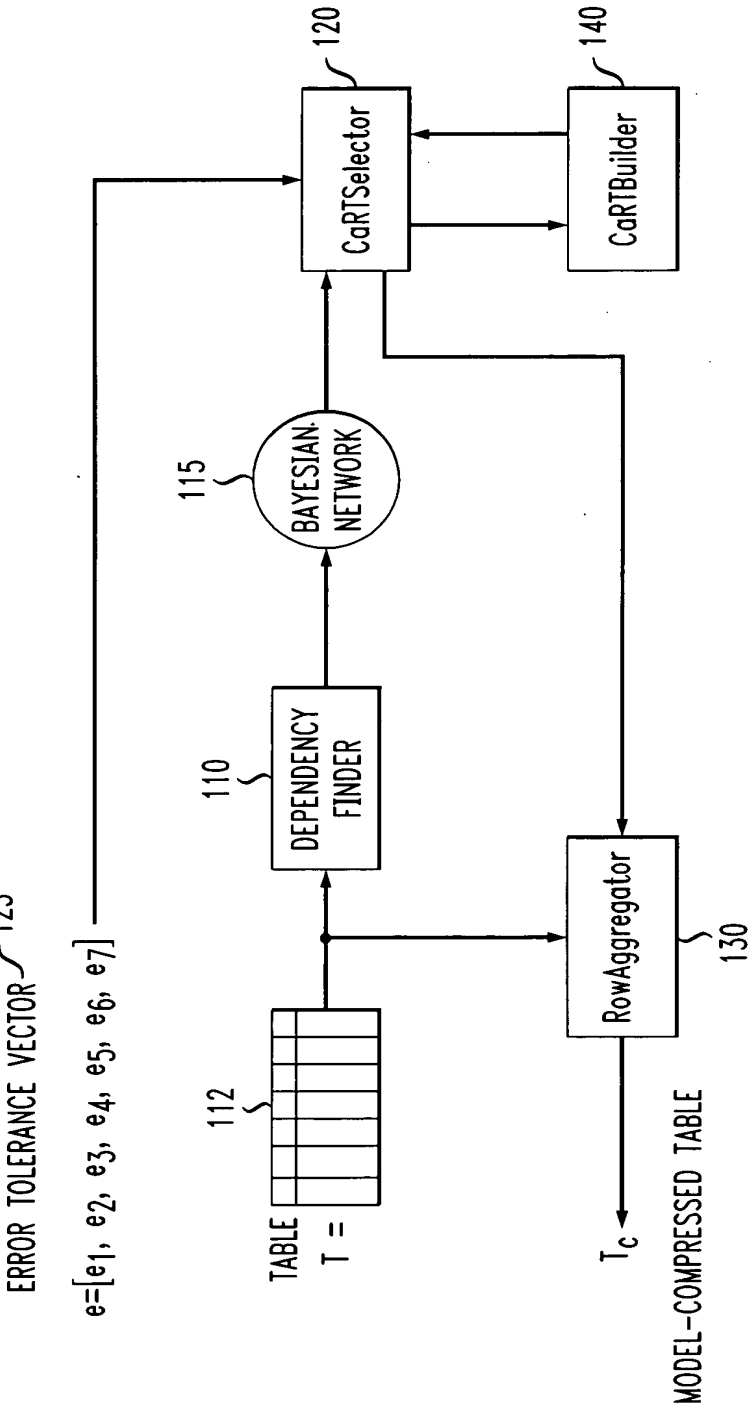


FIG. 2

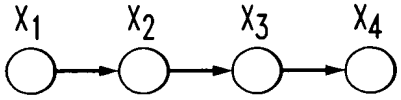
The Greedy CaRT Selection Algorithm

procedure Greedy $(T(X), \bar{e}, G, \theta)$
Input: n-attribute table T and n-vector of error tolerances $\bar{e}$; Bayesian network G on the set of attributes X and threshold θ on the relative benefit for selecting a CaRT predictor.
Output: A set of materialized (predicted) attributes X_{mat} ($X_{pred} = X - X_{mat}$) and a CaRT predictor for each $X_i \in X_{pred}$.

begin
1. $X_{mat} := X_{pred} := \Phi$
2. let $< X_1, X_2, \ldots, X_n >$ be the attributes in X sorted in *topological order* of G
3. for i :=1,...,n
4. if $\Pi(X_i) = \Phi$ then $X_{mat} := X_{mat} \cup \{X_i\}$ /* X_i must be materialized if it has no parents in G */
5. else
6. M := BuildCaRT $(X_{mat} \rightarrow X_i, e_i)$
7. if (MaterCost (X_i) / PredCost $(X_{mat} \rightarrow X_i) > \theta$) then $X_{pred} := X_{pred} \cup \{X_i\}$
8. else $X_{mat} := X_{mat} \cup \{X_i\}$
9. end
10. end

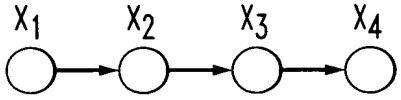
end

FIG. 3A



BAYESIAN NETWORK G

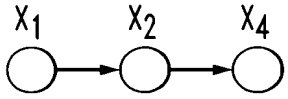
FIG. 3B



-infty 50 110 50

(b) Gtemp (FIRST ITERATION)

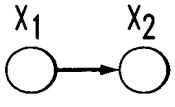
FIG. 3C



-infty -15 45

Gtemp (SECOND ITERATION)

FIG. 3D



-infty -60

Gtemp (THIRD ITERATION)

FIG. 4

The MaxIndependentSet CaRT Selection Algorithm

procedure MaxIndependentSet $(T(X), \bar{e}, G,$ neighborhood() $)$

Input: n-attribute table T and n-vector of error tolerances $\bar{e}$; Bayesian network G on the set of attributes X and function neighborhood () defining the "predictive neighborhood" of a node X_i in G (e.g., $\Pi(X_1)$ or $\beta(X_i)$).

Output: A set of materialized (predicted) attributes X_{mat} ($X_{pred} = X - X_{mat}$) and a CaRT predictor $PRED(X_i) \rightarrow X_i$ for each $X_i \in X_{pred}$.

begin

1. $X_{mat} := X$, $X_{pred} := \Phi$
2. $PRED(X_i) := \Phi$ for all $X_i \in X$, improve := true
3. while (improve $\neq$ false) do
4. for each $X_i \in X_{mat}$
5. mater_neighbors (X_i) := $(X_{mat} \cap \text{neighborhood}(X_i)) \cup \{PRED(X) : X \in \text{neighborhood}(X_i), X \in X_{pred}\} - \{X_i\}$
6. M := BuildCaRT (Mater_neighbors $(X_i) \rightarrow X_i$, e_i)
7. let $PRED(X_i) \subseteq$ mater_neighbors (X_i) be the set of predictor attributes used in M
8. $cost\_change_i := 0$
9. for each $X_j \in X_{pred}$ such that $X_i \in PRED(X_j)$
10. $NEW\_PRED_i(X_j) := PRED(X_j) - \{X_i\} \cup PRED(X_i)$
11. M :=BuildCaRT $(NEW\_PRED_i(X_j) \rightarrow X_j, e_j$
12. set $NEW\_PRED_i(X_j)$ to the (sub) set of predictor attributes used in M
13. $cost\_change_i := cost\_change_i$ + (PredCost (PRED $(X_j) \rightarrow X_j$) - PredCost $(NEW\_PRED_i(X_j) \rightarrow X_j)$)
14. end
15. end

FIG. 4 (cont)

```
16.      build an undirected, node-weighted graph G_temp = (X_mat,
         E_temp) on the current set of materialized
17.      attributes X_mat, where:
18.           (a) E_temp := { (X,Y) : for all pairs X, Y ∈X_pred } ∪
19.                                   { (X_i, Y) : for all Y ∈ X_mat }
20.           (b) weight (X_i) := MaterCost (X_i) -PredCost (PRED(X_i)
              → X_i) +cost_change_i  for each X_i ∈ X_mat
21.      S := FindWMIS (G_temp)     /* select (approximate) maximum
         weight independent set in G_temp
22.                                      as "maximum-benefit" subset of
                                         predicted attributes */
23.      if (Σ_{X∈S} weight (X) ≤ 0) then improve := false
24.      else/* update X_mat, X_pred , and the chosen CaRT predictors */
25.           for each X_j ∈ X_pred
26.                if (PRED (X_j) ∩ S = {X_i}) then PRED (X_j) :=
                     NEW_PRED_i (X_j)
27.           end
28.           X_mat := X_mat - S,  X_pred := X_pred ∪ S
29.      end
30.  end /* while */
end
```

FIG. 5

Algorithm for Estimating Lower Bound on Subtree Cost

```
procedure LowerBound      (N, e_i, b)
Input:  Leaf N for which lower bound on subtree cost is to be
        computed; error tolerance e_i for attribute X_i; bound b
        on the maximum number of internal nodes in subtree
        rooted at N.
Output: Lower bound L(N) on cost of subtree rooted at N.
begin
1.   for i := to r
2.        minOut [i, 0] :=i
3.   for J := 1 to b + 1
4.        minOut [0, j] :=0
5.   1 :=0
6.   for i := 1 to r
7.        while x_i - x_1+1 > 2_ei
8.        1 :=1 = 1
9.    for j := 1 to b + 1
10.       minOut [i,j] := min {minOut[i - 1,j] + 1, minOut [1,j-1]
11.  end
12.  L(N) := ∞
13.  for J := 0 to b
14.       L(N) :=min {L(N) , 2j + 1 + j log (|X_i|) + (j + 1 + minOut
          (r, j+1)) log (|dom(X_i)|)}
15.  L(N) := min {L(N) , 2b + 3 + (b + 1) log (|X_i|) + (b +2) log
     (|dom(X_i) |) }
16.  return L (N)
end
```

FIG. 6

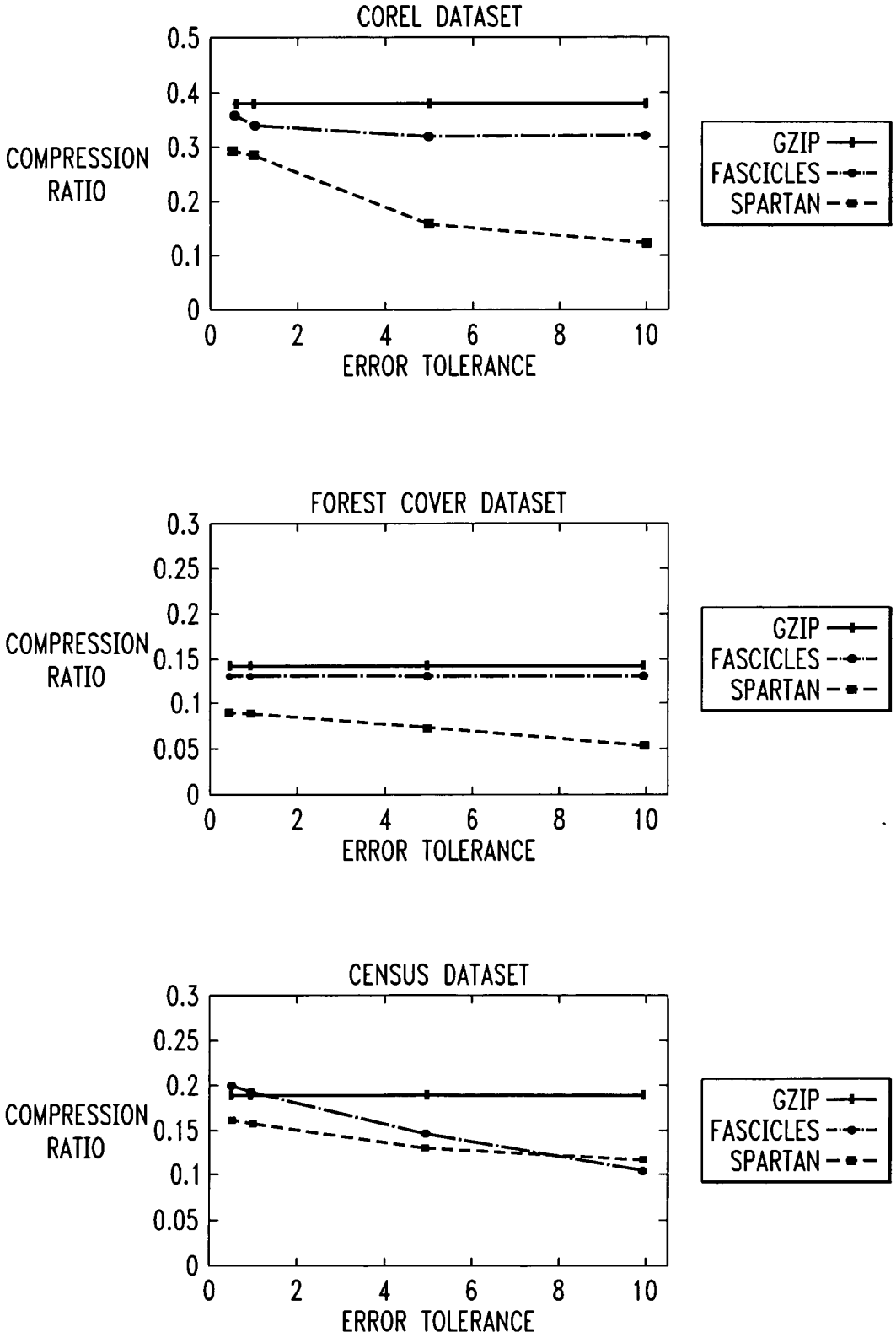


FIG. 7A

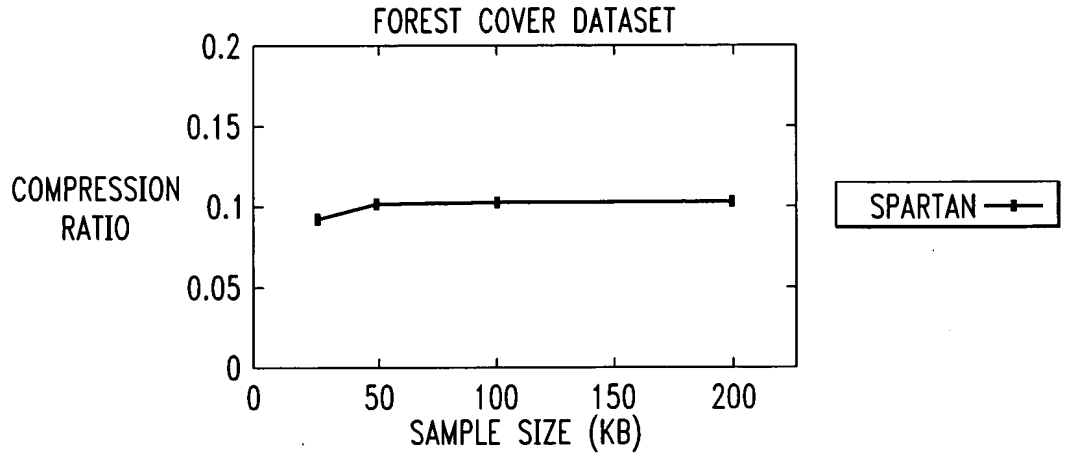


FIG. 7B

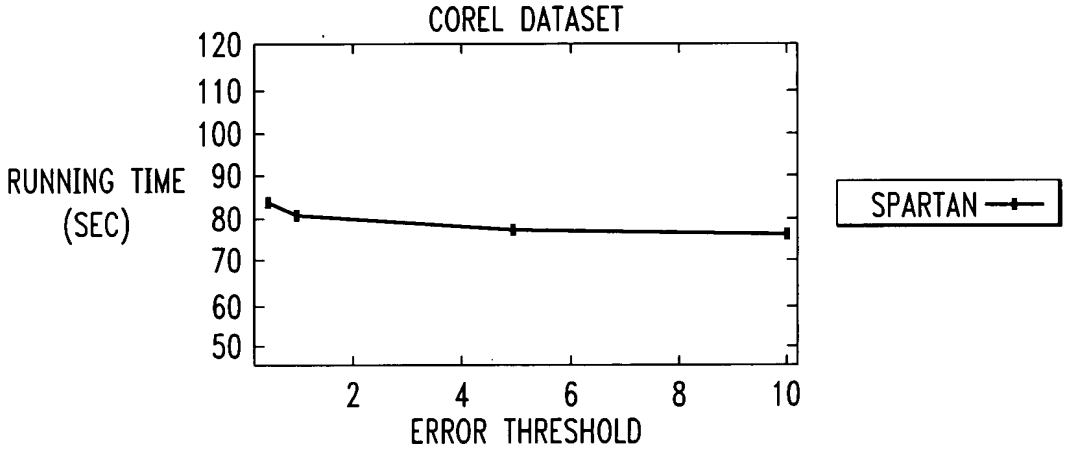


FIG. 7C